Data profiling project - Saheli Kar

Part 1:

I am choosing a data set of Real Estate. The csv file is shared in the GitHub.

RealEstateDf<-data1<- read.csv(file.choose(), header=TRUE)
head(RealEstateDf)</pre>

```
zip state beds baths
##
                               city
                 street
sq__ft
              type
## 1
                                                    2
           3526 HIGH ST SACRAMENTO 95838
                                                           1
                                              CA
836 Residential
            51 OMAHA CT SACRAMENTO 95823
                                              CA
                                                    3
                                                           1
1167 Residential
         2796 BRANCH ST SACRAMENTO 95815
                                                    2
                                                           1
                                              CA
796 Residential
       2805 JANETTE WAY SACRAMENTO 95815
                                              CA
                                                    2
                                                           1
852 Residential
        6001 MCMAHON DR SACRAMENTO 95824
                                              CA
                                                    2
                                                           1
797 Residential
                                                    3
## 6 5828 PEPPERMILL CT SACRAMENTO 95841
                                              CA
                                                           1
1122
           Condo
##
                         sale_date price latitude longitude
## 1 Wed May 21 00:00:00 EDT 2008 59222
                                             38.63
                                                       -121.4
## 2 Wed May 21 00:00:00 EDT 2008 68212
                                             38.48
                                                       -121.4
## 3 Wed May 21 00:00:00 EDT 2008 68880
                                             38.62
                                                       -121.4
## 4 Wed May 21 00:00:00 EDT 2008 69307
                                             38.62
                                                       -121.4
## 5 Wed May 21 00:00:00 EDT 2008 81900
                                             38.52
                                                       -121.4
## 6 Wed May 21 00:00:00 EDT 2008 89921
                                                       -121.3
                                             38.66
```

Part 2:

The variables in the Sacramento Real Estate data sets are:

```
street - The street address of the property.
```

city - The city where the property is located.

zip - The zip code of the property.

state - The state where the property is located.

beds - The number of bedrooms in the property.

baths - The number of bathrooms in the property.

sq ft - The area of the property in square feet.

type - The type of the property whether it is Residential/Condominium/ Multifamily

sale_date - The date when the property is sold.
price - The price of the propert sold.
latitude - The latitude loaction of the property.
longitude - The longitude location of the property.

The correaltion of the variables are provided with ggpairs()

```
#install.package(psych)
#require(psych)
#pairs(~zip+beds+baths+sq__ft+type+sale_date+price,
data=RealEstateDf)
install.packages("GGally")
```

```
## Installing package into
'C:/Users/sonatushi/Documents/R/win-library/3.1'
## (as 'lib' is unspecified)
```

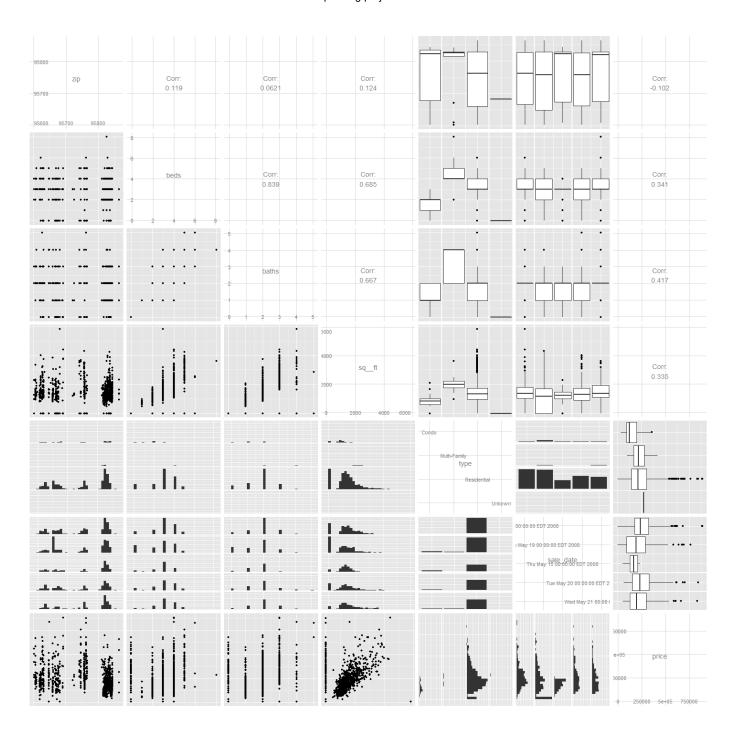
Error: trying to use CRAN without setting a mirror

```
require(GGally)
```

```
## Loading required package: GGally
```

```
## Warning: package 'GGally' was built under R version 3.1.1
```

```
ggpairs(data=RealEstateDf, columns=c("zip", "beds", "baths",
"sq__ft", "type", "sale_date", "price"))
```



Above we can see the correlation between the variables: zip, beds, baths, sq_ft, type, sale_date, price

Interesting correlation to be noted:

- Most of the 2 and 4 bedrooms property are sold in the areas around the zipcode 95800. Most of the 3 bedrooms are sold in the area with zip code 95600. Only one area had 8 bedrooms on sale.
- Very intersting positive correlation between beds and baths found.
- Also very positive correlation found between sqft and baths; and sqft and beds
- Most of the inexpensive property sold in the area around 95800. While Other areas sold realtively more exepnsive properties than the area under 95800.

- Most of the properties sold are Residential and most properties are inexpensive.
- Price and sq ft doesnt hold a spreaded correlation.

Recoding and binning:

Recoding the price and the sq__ft variable.

```
RealEstateDf$priceCategory[RealEstateDf$price<=100000]<-
"Inexpensive"
RealEstateDf$priceCategory[RealEstateDf$price>100000 &
RealEstateDf$price<=200000]<-"AveragePriced"
RealEstateDf$priceCategory[RealEstateDf$price>200000]<-
"Expensive"

RealEstateDf$AreaCategory[RealEstateDf$sq__ft<1000]<-
"Smaller"
RealEstateDf$AreaCategory[RealEstateDf$sq__ft>1000 &
RealEstateDf$sq__ft<2000] <-"Bigger"
RealEstateDf$AreaCategory[RealEstateDf$sq__ft>2000]<-"Very Big"

head(RealEstateDf, n=50)
```

##			reet city	zip
state ## 1	beds l		H ST SACRAMENTO	95838
CA	2	1		
## 2 CA	3	51 OMAH,	A CT SACRAMENTO	95823
## 3		2796 BRANCI	H ST SACRAMENTO	95815
CA ## 4	2	1 2805 JANETTE	WAY SACRAMENTO	95815
CA	2	1		
## 5 CA	2	6001 MCMAHOI	N DR SACRAMENTO	95824
## 6	_	5828 PEPPERMIL	L CT SACRAMENTO	95841
CA ## 7	3	1 6048 OGDEN NASH	WAY SACRAMENTO	95842
CA	3	2		
## 8 CA	3	2561 19TH	AVE SACRAMENTO	95820
## 9	11150	TRINITY RIVER DR Unit	114 RANCHO CORDOVA	95670
CA ## 10	2	7325 10TI	H ST RIO LINDA	95673
CA	3	2		
## 11 CA	3	645 MORRISON 2	AVE SACRAMENTO	95838
## 12	2	4085 FAWN	CIR SACRAMENTO	95823
CA ## 13	3	2 2930 LA ROS	A RD SACRAMENTO	95815
CA	1	1		

## 1 <i>1</i>		2112 KIDK WAY	SACRAMENTO	05022
## 14 CA 3	1		SACRAMENTO	93622
## 15		4533 LOCH HAVEN WAY	SACRAMENTO	95842
CA 2 ## 16	2	7340 HAMDEN PL	SACRAMENTO	95842
CA 2	2	7310 HAMBEN TE	SACIONEITIO	33012
## 17	1	6715 6TH ST	RIO LINDA	95673
CA 2 ## 18	Τ.	6236 LONGFORD DR Unit 1	CITRUS HEIGHTS	95621
CA 2	1			
## 19 CA 2	1		SACRAMENTO	95833
## 20		113 LEEWILL AVE	RIO LINDA	95673
CA 3	2		CITBUS USTCUTS	05621
## 21 CA 3	2	6118 STONEHAND AVE	CITRUS HEIGHTS	93021
## 22		4882 BANDALIN WAY	SACRAMENTO	95823
CA 4	2	7511 NAKVALE CT	NORTH HIGHLANDS	95660
CA 4	2			
## 24	2	9 PASTURE CT	SACRAMENTO	95834
CA 3	2	3729 BAINBRIDGE DR	NORTH HIGHLANDS	95660
CA 3	2	2020		05043
## 26 CA 3	2	3828 BLACKFOOT WAY	ANTELOPE	95843
## 27		4108 NORTON WAY	SACRAMENTO	95820
CA 3	1	1469 JANRICK AVE	SACRAMENTO	05832
CA 3	2	1409 JANKICK AVE	SACRAMENTO	93032
## 29	2	9861 CULP WAY	SACRAMENTO	95827
CA 4	2	7825 CREEK VALLEY CIR	SACRAMENTO	95828
CA 3				
## 31 CA 2		LAGUNA OAKS DR Unit 140	ELK GROVE	95758
## 32			NORTH HIGHLANDS	95660
CA 3	2	3100 EXPLORER DR	CACDAMENTO	05027
## 33 CA 3	2		SACRAMENTO	93027
## 34	2	7944 DOMINION WAY	ELVERTA	95626
CA 3		LAGUNA OAKS DR Unit 162	FLK GROVE	95758
CA 2				
## 36 CA 3	2	3920 SHINING STAR DR	SACRAMENTO	95823
## 37		5031 CORVATE ST	NORTH HIGHLANDS	95660
CA 3	2		CACRAMENTO	05022
## 38 CA 4	2	7661 NIXOS WAY	SACRAMENTO	95823
## 39		7044 CARTHY WAY	SACRAMENTO	95828
CA 4	2	2442 LARKSPUR LN	SACRAMENTO	95825
CA 1				
## 41	4800	WESTLAKE PKWY Unit 2109	SACRAMENTO	95835

i	$C\Lambda$	2	2)								
	CA ## 42				2178	63RI) A\	/E	SA	CRAN	MENTO	95822
	CA ## 43	3	2		871	8 ELI	< W.	ΑY	Е	ELK (GROVE	95624
	CA ## 44	3	2		RIDG	EPOIN	NT [OR .		ANTE	ELOPE	95843
	CA ## 45	2	2	<u>)</u>	7315	KOAI	Α (CT NORTI	4 H	ITGHI	ANDS	95660
	CA ## 46	4	2	<u>)</u>		2 ER:						95833
	CA	4	1									
	## 47 CA	4	2)	L SUNI							95823
	## 48 CA	4	1		7420 /	ALIX	PKV	٧Y	SA	CRAN	4ENTO	95823
	## 49 CA	4	2		320 N	ATOM/	A WA	ΑY	SA	CRAN	ИENTO	95838
	## 50 CA	3	2	4431	GREE	N TRI	EE [OR .	SA	CRAN	MENTO	95823
	## latitu	sq_	_ft [*]	type	9				S	ale_	_date	price
	## 1		836	Residential	Wed	Мау	21	00:00:0	00	EDT	2008	59222
	38.63 ## 2	1	167	Residential	Wed	Мау	21	00:00:0	00	EDT	2008	68212
	38.48 ## 3		796	Residential	l wed	May	21	00:00:0	00	EDT	2008	68880
	38.62 ## 4		852	Residential	l wed	Mav	21	00:00:0	00	EDT	2008	69307
	38.62 ## 5			Residential		_						81900
	38.52 ## 6		122			_		00:00:0				89921
	38.66					-						
	## 7 38.68			Residential		-						90895
	## 8 38.54	1	177	Residential	Wed	May	21	00:00:0	00	EDT	2008	91002
	## 9 38.62	,	941	Condo) Wed	May	21	00:00:0	00	EDT	2008	94905
	## 10 38.70	1	146	Residential	Wed	May	21	00:00:0	00	EDT	2008	98937
	## 11		909	Residential	Wed	Мау	21	00:00:0	00	EDT	2008	100309
	38.64 ## 12	1	289	Residential	Wed	Мау	21	00:00:0	00	EDT	2008	106250
	38.47 ## 13		871	Residential	Wed	Мау	21	00:00:0	00	EDT	2008	106852
	38.62 ## 14	1	020	Residential	l wed	May	21	00:00:0	00	EDT	2008	107502
	38.48 ## 15			Residential		_						
	38.67 ## 16		134			_		00:00:0				
	38.70					_						
	## 17 38.69		044	Residential	wed	МdУ	Z T	00:00:0	JU	בטו	2008	TT2/03

```
795
                   Condo Wed May 21 00:00:00 EDT 2008 116250
## 18
38.68
         588 Residential Wed May 21 00:00:00 EDT 2008 120000
## 19
38.61
        1356 Residential Wed May 21 00:00:00 EDT 2008 121630
## 20
38.69
        1118 Residential Wed May 21 00:00:00 EDT 2008 122000
## 21
38.71
        1329 Residential Wed May 21 00:00:00 EDT 2008 122682
## 22
38.47
        1240 Residential Wed May 21 00:00:00 EDT 2008 123000
## 23
38.70
        1601 Residential Wed May 21 00:00:00 EDT 2008 124100
## 24
38.63
         901 Residential Wed May 21 00:00:00 EDT 2008 125000
## 25
38.70
        1088 Residential Wed May 21 00:00:00 EDT 2008 126640
## 26
38.71
## 27
         963 Residential Wed May 21 00:00:00 EDT 2008 127281
38.54
        1119 Residential Wed May 21 00:00:00 EDT 2008 129000
## 28
38.48
## 29
        1380 Residential Wed May 21 00:00:00 EDT 2008 131200
38.56
## 30
        1248 Residential Wed May 21 00:00:00 EDT 2008 132000
38.47
                   Condo Wed May 21 00:00:00 EDT 2008 133000
## 31
        1039
38.42
        1152 Residential Wed May 21 00:00:00 EDT 2008 134555
## 32
38.69
## 33
        1380 Residential Wed May 21 00:00:00 EDT 2008 136500
38.57
## 34
        1116 Residential Wed May 21 00:00:00 EDT 2008 138750
38.71
## 35
        1039
                   Condo Wed May 21 00:00:00 EDT 2008 141000
38.42
        1418 Residential Wed May 21 00:00:00 EDT 2008 146250
## 36
38.49
        1082 Residential Wed May 21 00:00:00 EDT 2008 147308
## 37
38.66
        1472 Residential Wed May 21 00:00:00 EDT 2008 148750
## 38
38.48
## 39
        1146 Residential Wed May 21 00:00:00 EDT 2008 149593
38.50
## 40
         760
                   Condo Wed May 21 00:00:00 EDT 2008 150000
38.59
                   Condo Wed May 21 00:00:00 EDT 2008 152000
## 41
        1304
38.66
        1207 Residential Wed May 21 00:00:00 EDT 2008 154000
## 42
38.49
        1056 Residential Wed May 21 00:00:00 EDT 2008 156896
## 43
38.42
        1043 Residential Wed May 21 00:00:00 EDT 2008 161250
## 44
38.72
        1587 Residential Wed May 21 00:00:00 EDT 2008 161500
## 45
```

```
38.70
        1120 Residential Wed May 21 00:00:00 EDT 2008 164000
## 46
38.61
## 47
        1580 Residential Wed May 21 00:00:00 EDT 2008 165000
38.45
## 48
        1955 Residential Wed May 21 00:00:00 EDT 2008 166357
38.49
        1656 Residential Wed May 21 00:00:00 EDT 2008 166357
## 49
38.64
## 50
        1477 Residential Wed May 21 00:00:00 EDT 2008 168000
38.50
##
      longitude priceCategory AreaCategory
## 1
          -121.4
                   Inexpensive
                                     Smaller
## 2
         -121.4
                   Inexpensive
                                      Bigger
         -121.4
## 3
                                     Smaller
                   Inexpensive
## 4
         -121.4
                                     Smaller
                   Inexpensive
## 5
         -121.4
                                     Smaller
                   Inexpensive
## 6
         -121.3
                                      Bigger
                   Inexpensive
## 7
         -121.4
                                      Bigger
                   Inexpensive
## 8
         -121.5
                                      Bigger
                   Inexpensive
## 9
         -121.3
                                     Smaller
                   Inexpensive
         -121.4
## 10
                   Inexpensive
                                      Bigger
         -121.5 AveragePriced
                                     Smaller
## 11
##
   12
         -121.5 AveragePriced
                                      Bigger
## 13
                                     Smaller
         -121.4 AveragePriced
## 14
         -121.5 AveragePriced
                                      Bigger
##
   15
         -121.4 AveragePriced
                                      Bigger
## 16
         -121.4 AveragePriced
                                      Bigger
         -121.5 AveragePriced
                                     Smaller
##
   17
## 18
         -121.3 AveragePriced
                                     Smaller
## 19
         -121.5 AveragePriced
                                     Smaller
## 20
         -121.5 AveragePriced
                                      Bigger
## 21
         -121.3 AveragePriced
                                      Bigger
   22
##
         -121.4 AveragePriced
                                      Bigger
## 23
         -121.4 AveragePriced
                                      Bigger
## 24
                                      Bigger
         -121.5 AveragePriced
## 25
                                     Smaller
         -121.4 AveragePriced
## 26
         -121.4 AveragePriced
                                      Bigger
##
   27
         -121.5 AveragePriced
                                     Smaller
## 28
         -121.5 AveragePriced
                                      Bigger
##
   29
         -121.3 AveragePriced
                                      Bigger
## 30
         -121.4 AveragePriced
                                      Bigger
##
   31
         -121.4 AveragePriced
                                      Bigger
##
         -121.4 AveragePriced
  32
                                      Bigger
## 33
         -121.3 AveragePriced
                                      Bigger
##
         -121.4 AveragePriced
   34
                                      Bigger
## 35
         -121.4 AveragePriced
                                      Bigger
##
   36
         -121.5 AveragePriced
                                      Bigger
## 37
         -121.4 AveragePriced
                                      Bigger
## 38
         -121.5 AveragePriced
                                      Bigger
## 39
         -121.4 AveragePriced
                                      Bigger
                                     Smaller
## 40
         -121.4 AveragePriced
## 41
         -121.5 AveragePriced
                                      Bigger
## 42
         -121.5 AveragePriced
                                      Bigger
##
         -121.4 AveragePriced
  43
                                      Bigger
```

```
## 44
         -121.3 AveragePriced
                                     Bigger
## 45
         -121.4 AveragePriced
                                     Bigger
         -121.5 AveragePriced
## 46
                                     Bigger
## 47
         -121.4 AveragePriced
                                     Bigger
## 48
         -121.5 AveragePriced
                                     Bigger
## 49
         -121.4 AveragePriced
                                     Bigger
## 50
         -121.5 AveragePriced
                                     Bigger
```

Frequency of the Price Category with respect to the cities

```
require(dplyr)
## Loading required package: dplyr
## Warning: package 'dplyr' was built under R version 3.1.1
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##
       filter, lag
##
## The following objects are masked from 'package:base':
##
       intersect, setdiff, setequal, union
##
priceCategoryFreg<-select(RealEstateDf, city, priceCategory)</pre>
%>%
group_by(city, priceCategory) %>%
summarise(count=n())
show(priceCategoryFreq)
```

```
## Source: local data frame [64 \times 3]
## Groups: city
##
##
                  city priceCategory count
## 1
              ANTELOPE AveragePriced
                                          12
## 2
                                          21
              ANTELOPE
                            Expensive
## 3
                                           5
2
7
5
                AUBURN
                            Expensive
## 4
         CAMERON PARK AveragePriced
## 5
         CAMERON PARK
                            Expensive
## 6
            CARMICHAEL AveragePriced
                                          15
## 7
            CARMICHAEL
                            Expensive
## 8
                                          20
       CITRUS HEIGHTS AveragePriced
## 9
                                          13
       CITRUS HEIGHTS
                            Expensive
```

## 10	CITRUS HEIGHTS	Inexpensive	2
			1
	COOL	Expensive	1
## 12	DIAMOND SPRINGS	Expensive	
## 13	EL DORADO	Expensive	2 1 1 2 23
## 14	EL DORADO HILLS	Expensive	23
## 15	ELK GROVE	AvoragoBricod	19
		AveragePriced	
## 16	ELK GROVE	Expensive	92
## 17	ELK GROVE	Inexpensive	3
## 18	ELVERTA	AveragePriced	4
## 19		AveragePriced	1
		Averagerificed	7
## 20	FAIR OAKS	Expensive	3 4 4 5 1
## 21		AveragePriced	
## 22	FOLSOM	Expensive	16
## 23	FORESTHILL	AveragePriced	1
## 24	GALT	AveragePriced	9
## 25		Expansiva	13
	GALT	Expensive	
## 26	GARDEN VALLEY	Expensive	Ť
## 27	GOLD RIVER	Expensive	4
## 28	GRANITE BAY	Expensive	3
## 29	GREENWOOD	Expensive	1 4 3 1 3
## 30	LINCOLN		3
		AveragePriced	10
## 31	LINCOLN	Expensive	19
## 32	LINCOLN	Inexpensive	50
## 33	LOOMIS	Expensive	2
## 34	MATHER	Expensive	1
## 35	MEADOW VISTA	Expensive	2 1 1
## 36	NORTH HIGHLANDS	AveragePriced	18
## 37	NORTH HIGHLANDS	Expensive	2 1 2 9 1
## 38	NORTH HIGHLANDS	Inexpensive	1
## 39	ORANGEVALE	AveragePriced	2
## 40	ORANGEVALE	Expensive	9
## 41	PENRYN	Expensive	1
## 42	PLACERVILLE	Expensive	10
## 43	POLLOCK PINES	AveragePriced	1
## 44	POLLOCK PINES	Expensive	2
## 45	RANCHO CORDOVA		8
## 46	RANCHO CORDOVA	Expensive	19
## 47	RANCHO CORDOVA		1
			<u> </u>
_	RANCHO MURIETA	Expensive	2 1
## 49	RANCHO MURIETA	Inexpensive	Ţ
## 50	RIO LINDA	AveragePriced	6
## 51	RIO LINDA	Expensive	5
## 52	RIO LINDA	Inexpensive	3
## 53	ROCKLIN	Expensive	6 5 3 17
	ROSEVILLE	AveragePriced	10
## 55	ROSEVILLE	Expensive	38
## 56	SACRAMENTO	AveragePriced	200
## 57	SACRAMENTO	Expensive	182
## 58	SACRAMENTO	Inexpensive	70
## 59	SHINGLE SPRINGS	Expensive	í
## 60			1
	SLOUGHHOUSE	Inexpensive	
## 61	WALNUT GROVE	Expensive	Ţ
## 62	WEST SACRAMENTO	AveragePriced	2
## 63	WEST SACRAMENTO	Expensive	1 2 1 5
## 64	WILTON	Expensive	5
			_

Frequency of the Area Category with respect to the cities

```
require(dplyr)
areaCategoryFreq<-select(RealEstateDf, city, AreaCategory)
%>%
group_by(city,AreaCategory) %>%
summarise(count=n())
show(areaCategoryFreq)
```

```
Source: local data frame [79 x 3]
## Groups: city
##
##
                   city AreaCategory count
## 1
              ANTELOPE
                               Bigger
                                          22
## 2
                                            2
                              Smaller
              ANTELOPE
## 3
                                            9
1
              ANTELOPE
                             Very Big
## 4
                               Bigger
                AUBURN
                                            3
## 5
                              Smaller
                AUBURN
                                            ī
## 6
                             Very Big
                AUBURN
                                            8
## 7
          CAMERON PARK
                              Smaller
## 8
                                            1
          CAMERON PARK
                             Very Big
## 9
                                          14
            CARMICHAEL
                               Bigger
## 10
                                            2
3
                              Smaller
            CARMICHAEL
## 11
                             Very Big
            CARMICHAEL
                                            1
## 12
            CARMICHAEL
                                           31
##
   13
       CITRUS HEIGHTS
                               Bigger
                                            3
## 14
                              Smaller
       CITRUS HEIGHTS
                                           1
##
   15
       CITRUS HEIGHTS
                             Very Big
                                           1
## 16
                   COOL
                               Bigger
##
   17 DIAMOND SPRINGS
                                            1
                               Bigger
                                            2
2
##
   18
             EL DORADO
                               Bigger
## 19 EL DORADO HILLS
                               Bigger
   20 EL DORADO HILLS
                              Smaller
                                          20
  21 EL DORADO HILLS
                                           1
                             Very Big
##
   22
                                          61
             ELK GROVE
                               Bigger
## 23
                              Smaller
                                          10
             ELK GROVE
## 24
             ELK GROVE
                             Very Big
                                          43
## 25
                                            4
               ELVERTA
                               Bigger
                                            5
2
7
2
8
## 26
                               Bigger
             FAIR OAKS
                              Smaller
## 27
             FAIR OAKS
## 28
             FAIR OAKS
                             Very Big
## 29
                FOLSOM
                               Bigger
                              Smaller
## 30
                FOLSOM
## 31
                FOLSOM
                             Very Big
                                            1
##
  32
                              Smaller
            FORESTHILL
## 33
                                          17
                   GALT
                               Bigger
   34
##
                   GALT
                              Smaller
                                            1
## 35
                   GALT
                             Very Big
                                            4
```

```
36
         GARDEN VALLEY
                               Smaller
                                             1
                                             2
   37
            GOLD RIVER
                                Bigger
                               Smaller
   38
                                             1
            GOLD RIVER
##
   39
                                             1
            GOLD RIVER
                             Very Big
##
           GRANITE BAY
                             Very Big
  40
##
   41
             GREENWOOD
                              Very Big
##
                                             5
   42
                LINCOLN
                                Bigger
                               Smaller
                                           66
##
   43
                LINCOLN
##
                                             1
   44
                LINCOLN
                              Very Big
                                             2
##
   45
                 LOOMIS
                                Bigger
##
                                             1
   46
                 MATHER
                             Very Big
                                             1
##
   47
                                Bigger
          MEADOW VISTA
                                           16
   48 NORTH HIGHLANDS
                                Bigger
##
   49
                               Smaller
                                             5
      NORTH HIGHLANDS
                                             9
##
   50
            ORANGEVALE
                                Bigger
##
  51
            ORANGEVALE
                              Very Big
##
  52
                                             1
                 PENRYN
                                Bigger
                                             2
##
   53
           PLACERVILLE
                                Bigger
##
                               Smaller
   54
           PLACERVILLE
                                             2
##
   55
           PLACERVILLE
                              Very Big
##
   56
                                             3
         POLLOCK PINES
                                Bigger
                                           11
##
   57
                                Bigger
        RANCHO CORDOVA
##
                                             9
   58
        RANCHO CORDOVA
                               Smaller
##
   59
        RANCHO CORDOVA
                             Very Big
                                             1
##
   60
        RANCHO CORDOVA
##
   61
        RANCHO MURIETA
                               Smaller
                                             1
                                             2
##
   62
        RANCHO MURIETA
                             Very Big
                                           11
##
   63
             RIO LINDA
                                Bigger
                               Smaller
##
   64
                                             3
             RIO LINDA
                                            6
##
                                Bigger
   65
                ROCKLIN
                               Smaller
                                            9
##
   66
                ROCKLIN
##
   67
                             Very Big
                ROCKLIN
   68
##
             ROSEVILLE
                                Bigger
                                           14
                               Smaller
##
   69
                                           24
             ROSEVILLE
##
   70
             ROSEVILLE
                              Very Big
                                           10
##
                                          308
   71
            SACRAMENTO
                                Bigger
##
                               Smaller
   72
                                           94
            SACRAMENTO
##
                             Very Big
                                           50
            SACRAMENTO
##
   74
                               Smaller
                                             1
      SHINGLE SPRINGS
##
   75
           SLOUGHHOUSE
                             Very Big
                                             1
##
   76
                                Bigger
          WALNUT GROVE
                               Smaller
      WEST SACRAMENTO
##
   78
                               Smaller
                                             1
                 WILTON
## 79
                 WILTON
                             Very Big
```

Distribution of property prices with respect to longitude and latitude

require(ggplot2)

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.1.1
```

require(reshape)

```
## Loading required package: reshape
```

require(gridExtra)

```
## Loading required package: gridExtra
```

```
## Warning: package 'gridExtra' was built under R version 3.1.1
```

Loading required package: grid

```
longitudeMap<-ggplot(data = RealEstateDf) +
geom_point(aes(x=longitude, y=priceCategory ))+
ggtitle("Distribution of property prices with respect to
longitude")

latitudeMap<-ggplot(data = RealEstateDf) +
geom_point(aes(x=latitude, y=priceCategory ))+
ggtitle("Distribution of property prices with respect to
latitude")

grid.arrange(longitudeMap, latitudeMap, ncol=1 )</pre>
```

